DOCUMENT RESUME

ED 153 255 CS 204 095

TITLE Testing and Evaluation in Reading: Abstracts of Doctoral Dissertations Published in "Dissertation Abstracts International," March through December 1977 (Vol. 37 No. 9 through Vol. 38 No. 6).

INSTITUTION ERIC Clearinghouse on Reading and Communication Skills, Urbana, Ill.

PUB DATE 78

NOTE 11p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.

DESCRIPTORS *Annotated Bibliographies; *Cloze Procedure; Doctoral Theses; Elementary Secondary Education; Informal Reading Inventory; *Readability Formulas; Reading Comprehension; *Reading Research; Reading Skills; *Reading Tests; Student Attitudes; Test Reliability; Test Validity; Word Lists

IDENTIFIERS Right to Read; *Teacher Knowledge

ABSTRACT

This collection of abstracts is part of a continuing series providing information on recent doctoral dissertations. The 15 titles include studies of an informal reading inventory; an item-type to test phonic and context skills; reliability of residual gain scores based on means for a standardized reading achievement measure; alternatives to reading expectancy formulas; the cloze procedure; a test to determine reading comprehension of low socioeconomic status black sixth grade students; relationships between prelinguistic, linguistic, and psycholinguistic measures of readability; a behaviorally oriented reading attitude measure; a test of teacher knowledge of reading; measures for evaluating Colorado's Right-to-Read program; structural characteristics and item difficulty in standardized reading comprehension tests; and reading level based on graded word lists versus the Durrell Analysis of Reading Difficulty. (AA)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED153255

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

# Testing and Evaluation in Reading:

Abstracts of Doctoral Dissertations Published in Dissertation Abstracts International, March through December 1977 (Vol. 37 No. 9 through Vol. 38 No. 6)

Compiled by the Staff of

ERIC/RCS

ERIC Clearinghouse on Reading and Communication Skills
Urbana, Illinois

CS 204095

The dissertation titles contained here are published
with permission of the University Microfilms International,
publishers of Dissertation Abstracts International
(copyright ©1977 by University Microfilms International),
and may not be reproduced without their prior permission.

This bibliography has been compiled as part of a continuing series designed to make available information on relevant dissertations to users of the ERIC system. Monthly issues of Dissertation Abstracts International are reviewed in order to compile abstracts of dissertations on related topics, which thus become accessible in searches of the ERIC data base. Ordering information for the dissertations themselves is included at the end of the bibliography.

Abstracts of the following dissertations are included in this collection:

Braswell, Charlie Nell Powers

A COMPARISON OF TWO METHODS OF ADMINISTERING AN INFORMAL READING INVENTORY

Garcia, Cara Lynne

THE SCORE PRECISION OF AN ITEM TYPE TO TEST PHONIC AND CONTEXT SKILLS SEPARATELY AND CONJOINTLY

Graves, Gordon Richard

LONGITUDINAL RELIABILITY OF RESIDUAL GAIN SCORES BASED ON MEANS FOR A STANDARDIZED READING ACHIEVEMENT (POSTTEST) MEASURE RELATIVE TO FOUR ARRANGEMENTS OF PREDICTOR VARIABLES IN A SAMPLE OF 52 ELEMENTARY SCHOOLS FROM ONE DISTRICT

Heinrichs, Audrey Stewart

ALTERNATIVES TO READING EXPECTANCY FORMULAS

Holcomb, Carol Ann

USE OF THE CLOZE PROCEDURE AS A CRITERION FOR MEASURING THE READABILITY OF SELECTED PATIENT EDUCATION MATERIALS FOR OLDER ADULTS

Jones, Leroy Vernell

THE DEVELOPMENT OF A TEST TO DETERMINE THE READING COMPREHENSION ABILITY OF LOW SES BLACK SIXTH GRADE STUDENTS

Lindberg, Margaret Ann

A DESCRIPTIVE ANALYSIS OF THE RELATIONSHIP BETWEEN SELECTED "PRE-LINGUISTIC," LINGUISTIC AND PSYCHOLINGUISTIC MEASURES OF READABILITY

Mikulecky, Larry Joseph

THE DEVELOPING, FIELD TESTING, AND INITIAL NORMING OF A SECONDARY/ADULT LEVEL READING ATTITUDE MEASURE THAT IS BEHAVIORALLY ORIENTED AND BASED ON KRATHWOHL'S TAXONOMY OF THE AFFECTIVE DOMAIN.

Narang, Harbans Lal

THE DEVELOPMENT AND VALIDATION OF A TEST OF TEACHER KNOWLEDGE OF READING AT THE SECONDARY LEVEL

Phillips, Catherine Anne

THE VALIDITY OF CLOZE CRITERIA SCORES FOR SPECIFIC REGIONAL AND DIALECT GROUPS

Power, Marian Elizabeth

AN INVESTIGATION OF MODIFYING THE ORDER OF DIFFICULTY OF CLOZE PASSAGES WITH HIGHER EDUCATION OPPORTUNITY PROGRAM AND SPECIAL ADMIT STUDENTS

Ratliff, Sylvia Pettitt

AN EVALUATION OF COLORADO RIGHT-TO-READ TESTING FORMAL AND INFORMAL MEASURES OF TEACHER ATTITUDE, NEEDS ASSESSMENT, TEACHER EVALUATION OF INSERVICE AND AWARENESS OF RESOURCES

Sethaputra, Chariya

THE EFFECT OF STRUCTURAL CHARACTERISTICS UPON ITEM DIFFICULTIES IN STANDARDIZED READING COMPREHENSION TESTS

Shibata, Frances Hatch

A TRANSFORMATIONAL ANALYSIS OF THE VALIDITY OF THE CLOZE PROCEDURE AS A MEASURE OF READABILITY

Veon, Betty Ann Spillman

A COMPARISON OF SELECTED GRADED WORD LISTS AND THE DURRELL ANALYSIS OF READING DIFFICULTY FOR READING LEVEL PLACEMENT

## A COMPARISON OF TWO METHODS OF ADMINISTERING AN INFORMAL READING INVENTORY

BRASWELL, Charlie Nell Powers, Ph.D.
University of South Carolina, 1976

Chairperson: Dr. Paul Berg

An equal number of boys and girls from second and sixth grades were tested by two methods of administering an informal reading inventory. All sixty-four subjects were tested by both methods.

Subjects were matched by grade, sex, academic ability and reading levels. The first subject of a matched pair was tested by one method first and then by the second method, but the second subject would receive the methods in reverse order.

The two methods selected for inclusion in this study were the Betts method (oral reading preceded by silent reading) and the McCracken method (oral reading of a passage and silent reading of another passage).

Four null hypotheses were proposed with the following results. I. There is no significant difference in male or female's instructional reading level scores. This null hypothesis failed to be rejected at the $p \leq .05$ (.0514). II. Second or sixth graders do not score significantly higher on their instructional reading level scores. This null hypothesis was rejected (.0002). This rejection was predictable as the subjects were within the normal range of academic ability and were progressing normally in reading. III. There is no significant difference in the instructional reading level of subjects regardless of the order of administration. This null hypothesis failed to be rejected (.5603). IV. The Betts method does not result in a significantly higher instructional level than the McCracken method. This null hypothesis was rejected (.0002). The two variables which moderated this rejection were sex of the subjects and order of administration.

Boys did equally well on both methods. Girls did as well on the Betts method but scored lower on the McCracken method. When the order of administration was Betts method first there was little difference in the subjects reading level scores, but if the McCracken method was administered first the McCracken score was lower.

Teachers and clinicians should be aware that the McCracken method appears to result in a lower instructional reading level for girls.

Order No. 77-13,880, 71 pages.

## THE SCORE PRECISION OF AN ITEM TYPE TO TEST PHONIC AND CONTEXT SKILLS SEPARATELY AND CONJOINTLY

GARCIA, Cara Lynne, Ph.D.
The University of Arizona, 1977

Director: Kenneth J. Smith

The purpose of this study was to determine the extent to which the item type employed in Subtests 1–5 of the TOPIC (Test of Phonics in Context) results in score precision. Score precision is comprized of (1) item precision for parallel items which make pass/fail decisions for knowledge of phonic elements, and (2) parallel subtest form reliability for use of decoding strategies: context plus phonics, phonics only, context only, or neither strategy. Item precision is a criterion-referenced aspect of the TOPIC while parallel subtest reliability is a norm-referenced aspect. The item type which was used to determine score precision was a short declarative sentence ending in a noun which was deleted except for the phonic element being tested. The sentence was read for the students who were then asked to choose one of four pictured nouns to complete the statement. For example: Mother let the cat out of the h___. Pictures: house, hammer, cage, necklace.

Two alternate forms of the TOPIC were administered over two successive days to 338 native English speaking students in grades Kindergarten through two in the Flowing Wells School District in Tucson, Arizona. Item precision for knowledge of phonic elements was computed using Cohen's Kappa and Meredith's Error Discrepancy indices for each combination of grade level and phonic element and for grades K–2 combined. Parallel subtest reliability for use of reading strategy was computed using the Pearson product-moment correlation coefficient for each combination of grade level and strategy and for grades K–2 combined.

The study indicated that the item type was not a precise indicator for knowledge of phonic elements or use of reading strategy. The factors which were thought to influence the lack of score precision were: (1) flaws in the construction of some items, (2) the nature of the instructional program to which the subjects had been exposed, (3) restricted range of knowledge of phonic skills at the time of test administration, (4) weaknesses in procedures used to categorize items as precise or imprecise, and (5) the length of the test. Recommendations were made which included establishing the score precision of this item type at a point in time when there is more variance in knowledge of phonic skills and refining the procedures by which items are categorized as precise and imprecise.

Order No. 77-26,959, 91 pages.

## LONGITUDINAL RELIABILITY OF RESIDUAL GAIN SCORES BASED ON MEANS FOR A STANDARDIZED READING ACHIEVEMENT (POSTTEST) MEASURE RELATIVE TO FOUR ARRANGEMENTS OF PREDICTOR VARIABLES IN A SAMPLE OF 52 ELEMENTARY SCHOOLS FROM ONE DISTRICT

GRAVES, Gordon Richard, Ed.D.
University of Southern California, 1976

Chairman: Professor Michael

Problem. For a sample of 52 elementary schools in a central California district the initial purpose of this correlational investigation was to ascertain the longitudinal reliability of residual gain scores (RSG) based on means in a standardized reading achievement measure administered as a criterion posttest (Variable 1) to third-grade pupils during each of three consecutive years (May 1972, May 1973, and May 1974) for each of four arrangements of predictor variables (four prediction methods) that included (a) two hard-to-change measures alone — percentage of minority pupils in a school (Variable 2) and corresponding percentage of pupils receiving free lunches (Variable 3) and (b) Variables 2 and 3 plus a standardized reading achievement pretest measure administered to first-grade children also during May (Variable 4) for which three different sets of scores created from three different ways of of matching pupils in the first and third grades provided Variables 4a, 4b, and 4c (and thus three additional prediction methods). Two other purposes were to determine the congruence, or stability, of RGS across four prediction methods for one time point (May 1974) and to identify relative contributions of Variables 2, 3, and 4a or 4b or 4c to prediction of posttest scores (Variable 1). Used to evaluate the effectiveness of school programs, RGS were defined as differences between observed mean scores on a criterion measure and corresponding predicted scores.

Methodology. All third-grade pupils at a given school constituted the sampling unit. Besides coefficients of concordance (W) and product-moment correlation coefficients to indicate longitudinal reliability, or stability, four stepwise multiple regression analyses corresponding to four prediction methods for each consecutive year were calculated. In forecasting Variable 1, four prediction methods included these variables: (1) Variables 2 and 3, (2) Variables 2, 3, and 4a (mean score in pretest for first-grade pupils at the same school where a

posttest was given concurrently — no matching involved), (3) Variables 2, 3, and 4b (first-grade mean score of current third-grade class two years previously with no effort to match individual pupils because of uncontrollable mobility — an imperfect matching), and (4) Variables 2, 3, and 4c (first-grade mean score of only those third-grade children formerly in first grade at the same school — a perfect matching).

Findings. (1) For three sets of RGS over three consecutive years, four statistically significant W coefficients varied from .61 to .65 across four prediction methods. (2) Relative to these four prediction methods ranges in statistically significant coefficients of three pairings of RGS were, respectively .48 to .54, .44 to .55, .38 to .56, and .38 to .61. (3) Coefficients indicating congruence between sets of RGS associated with four prediction methods for May 1974 ranged from .88 to .97. (4) In multiple-regression analyses of variables in four prediction methods for each of three consecutive years statistically significant multiple correlation coefficients varied from .88 to .91 with Variables 2 or 3 always contributing the largest proportion of variance to predicting the criterion measure.

Conclusions. (1) Statistically significant but only modest indices of longitudinal reliability in RGS over a three-year period were attainable. (2) For one year, high degrees of congruence between sets of RGS were achieved across four prediction methods. (3) Hard-to-change variables offered great promise as valid predictors of reading performance.

Recommendations. (1) Caution should be exercised in using RGS to determine program effectiveness for more than one school year. (2) Evaluators should give serious consideration to employing hard-to-change predictor variables. (3) Additional multivariate research efforts should be initiated to determine reliability and validity of RGS over several consecutive years in diverse school communities.

## ALTERNATIVES TO READING EXPECTANCY FORMULAS

HEINRICHS, Audrey Stewart, Ph.D.
University of Illinois at Urbana-Champaign, 1976

In this era of controversy over IQ testing, it seemed valuable to devise a method of estimating reading expectancy that did not rely upon IQ scores. The research design was descriptive, a correlational study, intended to be hypothesis generating. These questions were examined: 1. Given a choice of four tests selected from the literature, which exclude IQ, what is the most efficient system for predicting silent reading achievement that can be devised using stepwise multiple regression analysis? 2. Of two prediction methods, one using IQ and one not, which is the more efficient in predicting silent reading of third and fourth grade children? 3. What are some of the characteristics of students who fall beyond one standard error of estimate from the prediction line using the IQ method as compared with characteristics of students who fall beyond one standard error of estimate from the prediction line using the method which excludes IQ?

The method that employed IQ scores in the determination of reading expectancy was exemplified by three systems: the Bond and Tinker, the Monroe, and the Horn reading expectancy formulas. For the non-IQ method, four tests were selected from the research literature as having potential for predicting reading achievement. Stepwise multiple regression analysis was applied.

All third and fourth grade students (N = 130) in a northern New England middle-income town were tested with the Stanford Achievement Tests, the Durrell Listening-Reading Series: Vocabulary Listening and Paragraph Listening, the Oral Directions subtest of the Detroit Tests of Learning Aptitude. IQ test scores were collected from the student files. Attrition reduced the sample to N = 68.

An initial test on the listening vocabulary and listening paragraph tests determined that the total score for the combined tests was more efficient in predicting reading achievement for this sample than was either test alone. Stepwise multiple regression analysis, applied to the total listening score, the arithmetic computation test, and the oral directions test selected the total listening score first ($F = 74.03$; $p < .001$) and arithmetic computation second ($F = 0.74$; $p > .39$). The addition of the second step was not significant ($p = .05$) and was rejected. Question 1 was answered by the equation: $\hat{Y} = -0.28 + (1.09 X$ Listening Total Score).

The correlations of the silent reading achievement test with the four predictive systems were compared, using a test of the significance of the difference between correlations on dependent samples. The non-IQ method described above correlated $R = .73$, the Bond and Tinker $R = .52$, the Monroe $R = .57$, and the Horn $R = .48$. The difference between the non-IQ regression equation above and the next highest correlation, the Monroe, was significant, $p < .05$. Question 2 was answered: of the two prediction methods, the best fit was obtained with the method that did not use IQ.

For Question 3, residuals were calculated; outliers were defined as those whose scores lay beyond one standard error of estimate from the regression line. Both methods showed a decided sex bias: a higher percentage of girls than boys achieved significantly above prediction; a higher percentage of boys than girls achieved significantly below prediction. Other comparisons among the outliers were made.

The conclusion was drawn that for this population, the total score of the listening sections of the Durrell Listening-Reading Series made a more accurate prediction of reading achievement on the Stanford Achievement Tests than did the Bond and Tinker, the Monroe, and the Horn reading expectancy formulas. The listening prediction system has the additional advantage of being independent of IQ scores. Suggestion was made that research be conducted on reasons for the sex differential. Additional research on dissimilar populations like Black or bi-dialect populations should be made for the purpose of contrasting those findings with these. Order No. 77-9023, 153 pages.

## USE OF THE CLOZE PROCEDURE AS A CRITERION FOR MEASURING THE READABILITY OF SELECTED PATIENT EDUCATION MATERIALS FOR OLDER ADULTS

HOLCOMB, Carol Ann, Ph.D.
Oregon State University, 1977

### Purpose

The major purpose of this study was to determine the readability of selected patient education materials for older adults using the cloze procedure as a criterion. Actual cloze test scores from three selected passages on hypertension were compared with the Coleman readability scores on those same passages. Age and the number of years of schooling completed for each subject were examined to determine their effect upon predicting the cloze test scores on each passage. Variations in the scores on the three selected passages were also examined.

### Procedure

The population consisted of adults, 60 years of age and over, who eat a noon meal at the 13 program sites operated by the Oregon District Four Elderly Nutrition Program. A random sample was drawn from a master list of all participants at the 13 sites. A total of 84 older adults comprised the sample for this study.

Each subject completed one cloze test in his/her own home with no time limit. The tests were scored on the basis of the number of exact completions (synonyms were not scored) for the 50 deletions. The raw scores were converted to percentage of correct completions for the purpose of analysis and

interpretation. Each subject also supplied information on his/her age and the number of years of schooling completed. The data for each of the cloze test forms were subjected to multivariate analysis, one-way analysis of variance and multiple comparisons analysis. The cell sizes for all three cloze test forms were equal. For all tests of hypotheses, findings for which the probability is less than .05 were reported as non-significant.

Findings

The major findings were: 1. Older adults made significantly ($p=<.01$) lower scores on the cloze tests than predicted by the Coleman readability formula. 2. Schooling correlated more significantly ($p=<.01$) than did age with the actual cloze test scores. 3. Schooling had a more significant ($p=<.01$) effect than did age upon predicting cloze test scores. 4. Cloze test, Form B, was significantly ($p=.05$) more difficult than cloze test, Form A.

Based on these findings, it was concluded that the patient education materials selected were "not very readable" for the older adults in the Oregon District Four Elderly Nutrition Program. Whereas all three passages tested were of approximately equal difficulty in terms of the language variables alone, the cloze procedure was able to distinguish differences in concept difficulty. Discussion of the findings included implications for the use of the cloze procedure by health educators for patient education of older adults and for use in evaluating the mental functioning of older adults. Suggestions for further research included: (1) the provision of additional variables, (2) replication of the study with additional patient education materials, (3) the sampling of other populations, (4) testing of the instruction sheet for readability, and (5) consistency in the use of the instruction sheet. Order No. 77-23,406, 128 pages.

## THE DEVELOPMENT OF A TEST TO DETERMINE THE READING COMPREHENSION ABILITY OF LOW SES BLACK SIXTH GRADE STUDENTS

JONES, Leroy Vernell, Ph.D.
University of California, Berkeley with
San Francisco State University, 1976

There have been a surprisingly small number of experimental studies regarding the nature of skills involved in reading comprehension, and there appear to be no valid means of determining the comprehension ability of low SES students. The purpose of this study was the development of a test which would more effectively measure the reading comprehension ability of low SES Black sixth-grade students. Test development was based on the premise that comprehension can best be determined when the test materials (paragraphs read) heighten interest and thus generate task attentiveness.

The new test instrument was a pictorial version of the Gilmore Oral Reading Test. The pictorial version included the same reading paragraphs, error checks, and comprehension questions as the standard version. The variation appeared in the manner in which students were asked to respond to the comprehension questions. Instead of requiring free recall, the pictorial version provided four visual selection possibilities for each question. Black subjects were depicted in the picture frames. Correct response was determined by the selection of the appropriate picture, name, or number.

The major hypothesis of the study was that low SES Black sixth-grade students would answer more reading comprehension questions correctly when administered the pictorial version of the Gilmore Oral Reading Test than when administered the standard version.

Sixty low SES Black sixth-grade students were assigned randomly to four groups and administered the pictorial and standard versions of the Gilmore Oral Reading Test in counterbalanced order. In addition, 30 students were administered a post-test to determine the reliability of the pictorial version, while 15 students were administered a multiple choice comprehension test (without pictures) to determine the significance of the pictures in the pictorial version as opposed to the multiple choice format. To obtain data concerning student interest, questions were developed to allow all students an opportunity to express a preference for either the standard or pictorial version, to evaluate the pictures, and to assess whether the pictures helped them answer the comprehension questions.

The major hypothesis was supported at the .001 level. It was postulated that students performed better because the pictures and the multiple choice format helped them determine the correct answers. These conclusions were supported by results of the student interest questionnaire, which indicated that an overwhelming majority of students preferred the pictorial version rather than the standard version. It was further postulated that student interest was of primary importance in determining better performance on the pictorial version. However, as the multiple choice comprehension test (without pictures) produced better scores than the pictorial version, additional research is needed to determine the impact of high interest materials which depict Black subjects in their content.

The findings of this study point to the need for development of appropriate, high interest materials to test the reading comprehension ability of low SES students.

Order No. 77-15,927, 177 pages.

## A DESCRIPTIVE ANALYSIS OF THE RELATIONSHIP BETWEEN SELECTED "PRE-LINGUISTIC," LINGUISTIC AND PSYCHOLINGUISTIC MEASURES OF READABILITY

LINDBERG, Margaret Ann, Ph.D.
Wayne State University, 1977

Major Professor: Kenneth Goodman

This study compared six readability measures with varied theoretical bases applied to five fourth grade basal reader stories, and evaluated the measures' potential usefulness to educators and publishers in the selection and use of instructional materials.

The "pre-linguistic" measures included the 1958 revision of the Dale-Chall and the Fry Readability Graph; the linguistic measures included Moir's Word-Content Ratio, and the Botel-Granowsky Syntactic Complexity Formula. The Cloze Readability Test and the Miscue Formative Depth Analysis comprised the two psycholinguistic measures of readability.

A related portion of this study probed the cloze procedure and miscue analysis as indicators of reading processing and comprehension by examining reader responses to both instruments using selected categories of the Goodman Taxonomy of Reading Miscues. Two taxonomic analyses were carried out: Cloze II (identical, non-identical, and omitted responses), and Cloze III (non-identical and omitted responses only).

The written responses of eight cloze subjects reading two passages taken from a single story (for a total of sixteen passages), and the first fifty oral miscues of sixteen readers reading the same story, were compared. All participants were considered to be average fourth-grade readers according to teacher judgment and standardized test scores.

Although the measures investigated in this study are based on theoretically disparate premises and procedures, the measures were generally able to place the stories within one grade level of fourth-grade--acceptable statistical parameters for readability measures.

That such diverse measures provided relatively similar data on levels of story difficulty may indicate that the measures are assessing broadly related phenomena.

However, none of the measures showed identical ranges or grade placements for the five stories, or specifically agreed with publisher judgment. Nor did the measures, for the most part, rank the stories in the same order of difficulty.

Two of the measures, the Fry and Dale-Chall, showed patterns of variability that seemed more related to the measurement characteristics of the instruments than to the language of the stories.

Since readability measures are usually expected to provide sequencing for publisher or classroom materials, the measure selected for use will alter the sequencing in meaningful ways. Additionally, grossly accurate information, which is all such measures can provide, is not useful for developing such sequencing or for planning instruction.

Only the two psycholinguistic measures, the Miscue Formative Depth Analysis and the cloze can be used to provide insightful information to teachers and publishers. The assumptions and procedures of the former, validated by this study, can provide in-depth conceptual, linguistic, and cultural data upon which to base the selection and use of instructional material.

The Cloze Readability Test, as evaluated by the Goodman Taxonomy, appears to be a useful predictor of readers' abilities to process the language of the text, and perhaps of reading comprehension as well.

It appears that the kinds of language-processing strategies used in oral reading, at least in non-conceptually dense materials, are reflected in both cloze and miscue processes. Cloze non-identical responses seemed to indicate a "miscueing process" similar to producing a non-expected oral response to print. Thus, there seems to be a universal process by which readers attempt, more or less successfully, to obtain meaning from print.

Bormuth's pre-1975 criteria seem more reflective of readability when compared to both oral reading and the other readability measures, than do his later criteria.

Finally, if the figures obtained in this study hold for larger populations, by doubling groups of readers' Cloze Readability Test Scores, publishers may have an easy means of estimating meaning-processing strategies of readers, as well as the predictive strength of the materials.

Order No. 77-23,992, 319 pages.

## THE DEVELOPING, FIELD TESTING, AND INITIAL NORMING OF A SECONDARY/ADULT LEVEL READING ATTITUDE MEASURE THAT IS BEHAVIORALLY ORIENTED AND BASED ON KRATHWOHL'S TAXONOMY OF THE AFFECTIVE DOMAIN

MIKULECKY, Larry Joseph, Ph.D.
The University of Wisconsin-Madison, 1976

Supervisor: Professor Kenneth L. Dulin

Test-oriented accountability seems to emphasize that which is testable. In order to emphasize the importance of reading-attitude, therefore, a sound, usable measure of reading-attitude is necessary. A review of relevant literature revealed a paucity of reading-attitude measures appropriate for mature readers.

The Mikulecky Behavioral Reading Attitude Measure was developed to be a sound reading-attitude measure appropriate for use with mature readers. To establish the instrument on sound theoretical foundations, all items were written with direct reference to the Hovland-Rosenberg tri-component model of attitude and to the stages of Krathwohl's Taxonomy of the Affective Domain. A pool of 40 items, each of which was designed to reflect a specific Krathwohl sub-stage, was reduced to 20 items after considering the evaluations of a panel of judges familiar with Krathwohl's Taxonomy and after an item analysis which eliminated all items that correlated at $r = .600$ or less with the sum of items reflecting the Krathwohl stage appropriate to each item. The hierarchical framework hypothesized by Krathwohl was supported by an analysis of subjects' item responses using a method for Scaling a Simplex developed by Henry Kaiser (Psychometrika, 1962). The MBRAM hierarchy gave evidence of a .933 out of a possible 1,000 goodness-of-fit to an ideal hierarchy. This was interpreted as empirical support for the Krathwohl theoretical foundation of the MBRAM.

A graduate-level seminar on Affective Domain measurement helped survey and refine all items to reflect everyday reading-related behaviors thereby establishing face validity. Correlations of concurrent validity ranging from .446 to .770 were established with such formal reading-attitude measures as the Estes Scale, the Dulin-Chester Scale, and the Kennedy-Halinski Reading Attitude Measure. The MBRAM correlated more highly with the Estes Scale and the Dulin-Chester Scale than either of those measures did with the other.

To establish construct validity, five informal criteria for reading-attitude (Self-reported Liking and Amount of Reading, Teacher and Classmate judgment of reading-attitude, and Number of books read in 6 months) were administered along with the MBRAM. All MBRAM correlations with these informal criteria were significant to the $p < .001$ level and the majority of correlations ranged from .500 to .791. The MBRAM correlated significantly more highly with these informal measures than did the other formal reading-attitude measures used in the study. Analysis of variance statistically demonstrated the ability of the MBRAM to discriminate subjects of high, average, and low reading-attitude as measured by the informal criteria.

The MBRAM demonstrated a test-retest reliability of .9116.

Correlation of MBRAM scores with nine different measures of academic and reading competency indicated that reading-attitude is related to competency, but separable from it. In most correlations, less than 10 percent of the variance between the MBRAM and competency measures was shared, and in some correlations, less than 1 per cent of the variance was shared.

The MBRAM was administered to 1750 subjects ranging from 7th grade through college-adult. 1,343 Of the subjects were public school students selected from urban, suburban, and rural populations. These subjects were randomly sampled to create a composite, stratified Wisconsin Population Model. Norms for the MBRAM are reported for each grade level in terms of this model and also in terms of urban, suburban, and rural populations. For ease of interpretation of scores, attitude-level scoring bands are provided. No significant differences in scores of urban, suburban, or rural subjects were found from 7th-10th grade, but rural subjects exhibited slightly higher MBRAM mean scores in the upper grades. Reading-attitude scores decreased slightly in all locations with each year in school. Order No. 76-28,162, 187 pages

## THE DEVELOPMENT AND VALIDATION OF A TEST OF TEACHER KNOWLEDGE OF READING AT THE SECONDARY LEVEL

NARANG, Harbans Lal, Ph.D.
University of Oregon, 1976

Adviser: Dr. Karl D. Hesse

The purpose of this study was to develop an instrument for measuring teacher knowledge of reading at the secondary level. A content outline was prepared from an examination of the literature on competency-based reading programs and methods textbooks in secondary reading. The topics selected on the basis of research evidence, expert opinion, and their practical use to the classroom teacher were organized in four areas as follows: I. GENERAL BACKGROUND (a). Reading and

Reading Problems. (b). Nature and Difficulty of Materials. II. READING SKILLS (a). Word Recognition and Vocabulary. (b). Comprehension. (c). Study Skills. III. INSTRUCTIONAL STRATEGIES (a). Motivational Techniques. (b) Lesson Plans and Study Guides. IV. MEASUREMENT AND EVALUATION (a). Reading Tests. (b). Informal Techniques. (c) Test Interpretation. The content outline provided a basis for the content validity of the test. Multiple-choice items were generated from this content and submitted to a panel of experts. The experts categorized the items according to the content outline, rated each item in terms of the importance of knowledge it measured, and answered each item. This was done to confirm the content validity and to prepare an answer key acceptable to the potential users of the test.

The preliminary test was administered to two groups, one consisting of secondary students and the other of teachers. An item analysis was carried out to determine the item difficulty, item validity, and reliability of the test. This analysis led to the rejection of some items and revision of others. The final version consisted of 45 items covering the four areas of the content and topics listed above.

The revised test was administered to 124 teachers and 64 students in secondary education. Their scores ranged from 11 to 40 with a mean of 24.5 and standard deviation of 6.3. The reliability coefficient obtained by KR-20 formula was .76.

The correlation between the four subtests of general background, reading skills, instructional strategies, and measurement and evaluation were low (from .2? to .37). This suggests that they measure different abilities. However, their reliability coefficients (from .35 to .58) were not high enough to warrant their use as independent subtests.

Teachers and students were compared on their performance and it was found that teachers achieved significantly higher than students. Persons who had taken a reading course did better than those who had not taken such a course. There was no significant difference between the mean scores of the junior and senior high school teachers. Performance on the basis of sex was also not significantly different. Similarily, subject-matter fields of the teachers and the amount of their experience did not affect their performance on this test.

On the basis of this study it was concluded that the test of teacher knowledge of reading at the secondary level is a valid and reliable instrument for measuring preservice and inservice teacher knowledge of reading.

Order No. 77-4745, 99 pages.

## THE VALIDITY OF CLOZE CRITERIA SCORES FOR SPECIFIC REGIONAL AND DIALECT GROUPS

PHILLIPS, Catherine Anne, Ph.D.
Arizona State University, 1977

This study began an examination of the applicability of currently accepted cloze criteria for populations of the southwest regions and dialect groups. The sample for the study consisted of 86 Black and 61 White (N = 147) seventh grade students selected from a middle school located in the southcentral section of Phoenix, Arizona. This section is classified as one consisting of a predominantly low socio-economic population.

The subjects were administered an appropriately graded cloze test and the reading comprehension subtest of the Comprehensive Test of Basic Skills. Their performance placements were ranked for each instrument. Employing cloze criteria levels currently being used, the subjects' cloze test placements and multiple-choice test placements were measured for differences. The cloze criteria levels then were manipulated several times in search of the least difference between frequency distributions. When the least difference was determined by chi-square analysis, the criterial score was identified. The Black subjects' cloze data were manipulated in search of a lesser difference between frequency distributions. Chi-square analyses were also employed for each of these manipulations. The White subjects' cloze data were manipulated using the criteria established by manipulation of the total group's scores and again using currently accepted criteria. The level of confidence selected for interpretation of the analyses was .05.

Statistical analyses of these manipulations resulted in the following findings: 1. There was a statistically significant difference between the subjects' cloze test placements and multiple-choice comprehension test placements when cloze criteria scores, as suggested by leading authorities, were applied. 2. The data were manipulated and measured in order to attain the least difference between frequency distributions. The least difference was found when the cloze criteria were adjusted as follows: 62% to 100% cloze test performance identified as independent reading level by leading authorities remained independent level; instruction level was adjusted from the range of 41% to 61%, which is suggested by leading authorities, to 45% to 61%; and frustration level performance was identified at 44% and below. There was no statistically significant difference as measured by chi-square analysis between the subjects' cloze test placements and multiple-choice comprehension test placements when the new cloze criteria were applied. 3. There was no statistically significant difference between seventh grade Black dialect subjects' cloze test placements and multiple-choice comprehension test placements when regionally adjusted cloze criteria were applied. 4. The White subjects' placements were measured for differences between frequency distributions using the regionally adjusted criteria and no significant difference was determined by chi-square analyses. Their placements were measured for differences between frequency distributions, using currently accepted criteria, and there was a significant difference as indicated by chi-square analysis.

The results of this study led to several recommendations which include the following: extension of this study with a sample to include Spanish-surname subjects selected from both seventh and eighth grades; extension of this study with instruments developed from the same reading selection; extension of this study with samples drawn from other areas of the Southwest regions of the United States; and extension of this study to include other socio-economic status groups.

Order No. 77-22,444, 100 pages.

## AN INVESTIGATION OF MODIFYING THE ORDER OF DIFFICULTY OF CLOZE PASSAGES WITH HIGHER EDUCATION OPPORTUNITY PROGRAM AND SPECIAL ADMIT STUDENTS

POWER, Marian Elizabeth, Ed.D.
Syracuse University, 1976

This study investigated the significance of gains in reading comprehension when using two methods of presenting cloze passages to college students who had been identified as being disadvantaged. The two methods of presentation consisted of administering cloze passages that had been evaluated in terms of readability levels in an easy-to-difficult hierarchical order and a difficult-to-easy hierarchical order of difficulty. This study also examined the effect of the graded college freshmen material on cloze passage performance.

The following null hypotheses were posited: $H_1$ There will be no significant difference in gains in reading comprehension among the following groups: a) HEOP and S.A. students who receive cloze passages in an easy-to-difficult readability order; b) HEOP and S.A. students who receive cloze passages in a difficult-to-easy readability order; c) HEOP and S.A. students who receive no cloze intervention. $H_2$ There will be no significant differences between treatment groups' mean scores on the graded cloze passages when varying the order of difficulty.

Twenty-two Higher Education Opportunity Program and 23 Special Admit students were randomly selected from a Developmental Reading Course in order to participate in this study. Assignment to two experimental groups and one control group were made by random numbers generated by computer. During the first meeting of classes all groups were pretested with the Diagnostic Reading Test (Triggs) prior to any reading instruction. Each day Experimental Groups I and II received the same lectures on reading and study skills as the Control Group III. Group I received 23 cloze passages constructed from college textbooks and evaluated in terms of readability levels, in an easy-to-difficult order. Group II received the passages in a difficult-to-easy hierarchical order. Groups I and II participated in small group discussions following the completion of the cloze exercises. Control Group III received no exercises or instruction in cloze. All groups met five days a week for 25 class sessions 1½ hours long. All subjects were retested with an alternate form of the DRT during the final meeting period.

A multivariate analysis of variance was utilized to test the first hypothesis. The resultant Wilks' lambda fell far short of significance at the .05 level, thus failing to reject the null hypothesis.

A two-way analysis of variance for testing significance of differences in mean scores between the two experimental groups indicated that there was a significant difference at the .05 level with the variance ratio F=4.24. The second null hypothesis was rejected.

The two independent variables of method of presentation were found to interact. Additional insight concerning the results of this study was obtained by computing tests of simple main effects. A significant difference was observed on cloze passage 9. Inspection of the cloze passages indicated that the mean score of cloze passage 9 was significantly higher in the difficult-to-easy order of presentation than its mean score in the easy-to-difficult order. The differences between Experimental Groups' I and II mean scores were not significant for passages 1 through 8 or 10 through 23.

Order No. 77-24,861, 154 pages.

## AN EVALUATION OF COLORADO RIGHT-TO-READ: TESTING FORMAL AND INFORMAL MEASURES OF TEACHER ATTITUDE, NEEDS ASSESSMENT, TEACHER EVALUATION OF INSERVICE AND AWARENESS OF RESOURCES

RATLIFF, Sylvia Pettitt, Ed.D.
University of Colorado at Boulder, 1977

Director: Professor Miles C. Olson

The major purposes of this study were to determine the following:

the effect of reading service education upon the attitudes of teachers

the perceptions of teachers regarding the effectiveness of inservice education

the extent of use of teacher needs assessments as the basis for inservice programs

director awareness of available resources, human and material, for implementing inservice education.

A review of the literature focused on elements of inservice education pertinent to this study: teacher change of behavior, evaluation, needs assessments and awareness of resources.

A series of questions was asked reflecting instruments used in each of the three steps in the evaluation: antecedent, transaction and outcome. Treatment of data was handled on two levels: descriptive and analytic. At the descriptive level, simple frequencies and percentages were computed and discussed descriptively. The Statistical Package for the Social Sciences (SPSS) provided simple means for post-inservice evaluation. At the analytic level, using the Statistical Package for the Social Sciences, chi-square tests were performed across 34 variables in order to answer questions concerning teacher and administrator attitudes toward inservice education.

Analysis of data produced the following major conclusions: 1. Teachers request inservice specific to their levels and needs. 2. Administrative support to the program will be given in various ways. 3. Parents will be supportive of this kind of program. 4. Classroom teachers may be effectively used to conduct inservice. 5. Lack of follow-up to inservice is a point of weakness in the program. 6. Inservice education appears to be perceived by teachers as a useful approach and should be continued.

In light of conclusions drawn, the following recommendations were made: 1. Data from the present study should be made available to upcoming programs. 2. Federal funding proposals should include funding for early identification of local district talent. 3. Professional persons conducting inservice should be aware of expressed need for useful inservice programs. 4. Follow-up should be an identified component of Right-to-Read programs and subject to evaluation. 5. Local districts should make a concentrated effort to provide multiple and varied inservice activities for secondary teachers. 6. Additional studies should be undertaken in order to stimulate improved evaluation of Title programs. 7. Inservice programs should include information and materials which allow for transfer to the classroom of concepts and skills.

Order No. 77-24,275, 172 pages.

## THE EFFECT OF STRUCTURAL CHARACTERISTICS UPON ITEM DIFFICULTIES IN STANDARDIZED READING COMPREHENSION TESTS

SETHAPUTRA, Chariya, Ph.D.
The Florida State University, 1977

Major Professor: F. J. King

The purpose of this study was to determine the effect of structural characteristics identified in reading passages and test items upon item difficulties in standardized reading comprehension tests. Structural characteristics investigated included the average word length, average sentence length, proportion of hard words outside the Dale list of 769 words, total number of words in each test item, human interest, reading interest and item types. The instruments used were the Comprehensive Tests of Basic Skills, Forms Q and S; the California Achievement Tests, Form A; the Sequential Tests of Educational Progress Series II, Form A; and the original Sequential Tests of Educational Progress, Form B. Empirical item difficulties provided by the publishers of these standardized tests were used as the dependent variable.

In general, it was found that the average word length, average sentence length and proportion of hard words outside the Dale list of 769 words identified in each reading passage and the average word length, total number of words and proportion of hard words identified in each test item have an influence upon item difficulty, whereas the human interest and reading interest variables failed to show significant effects. Item type was also shown to be a significant factor in determining the difficulty of a test item: inferential comprehension items were shown to be harder than literal comprehension items. A quadratic relationship was also found between the proportion of hard words identified in a test item and its difficulty.

The main conclusion of the study was that structural characteristics of reading passages associated with those in test items have an influence in predicting item difficulty in standardized reading comprehension tests.

Order No. 77-22,153, 104 pages.

## A TRANSFORMATIONAL ANALYSIS OF THE VALIDITY OF THE CLOZE PROCEDURE AS A MEASURE OF READABILITY

SHIBATA, Frances Hatch, Ph.D.
Syracuse University, 1976

The primary objective of this investigation is to determine how accurately cloze scores reflect the structural or transformational complexity of a passage and, therefore, the readability of that passage. Readability, or the difficulty a person has in comprehending a passage, was independently determined by a group of twenty-eight judges who ranked the passages from easy to hard.

The ability of transformational complexity to measure readability was tested by comparing the Transformational Complexity Score (TCS) for each passage with the rater rankings. The Transformational Complexity Scores used in the study were obtained by reconstructing the derivational history of each sentence and analyzing each derivation for the presence of twenty-nine transformations. The TCS is the total number of transformations in a passage. Using correlation analysis procedures, the ratio between rater rankings and the TCS was found to be significant. Not only did this confirm the assumption that transformational complexity is a valid index of reading difficulty, but also that the TCS was the best measure of readability employed in the investigation.

The mean cloze score for each passage was compared with the rater rankings and the TCS. The mean cloze score is the average number of incorrect restorations of the deletions in each passage. Twenty-five different cloze passages were constructed from five original selections by deleting every fifth word on a rotating basis. Each passage contained a total of forty deletions, and only exact restorations of the deleted words were counted as correct. When the cloze scores were compared with the rater rankings and the TCS, the correlation ratios were not significant. However, they did reflect trends in the rater rankings, suggesting that further studies of a nonlinear relationship should be undertaken.

In addition, the Dale-Chall grade level scores for each passage were calculated according to the Dale-Chall formula. When the scores were compared with each of the other measures of readability, none of the ratios was significant. Moreover, the extremely low correlation between the Dale-Chall scores and actual reader performance on the passages, as reflected by the cloze scores, casts doubt on the value of the grade level scores as predictors of reading difficulty. Nevertheless, the moderate correlation between the Dale-Chall and the Transformational Complexity Scores implies that a common demoninator is operative, an implication which deserves further study.

Although the results of the preceding comparisons fail to establish the TCS as a predictor of cloze performance, they provide a basis for additional research which examines the effects of redundancy and word order, as measured by the TCS and Dale-Chall scores, on cloze performance. Such experimental data would serve to strengthen the tentative conclusions of the present investigation while adding insight to the subject of readability and its role in the comprehension process. Order No. 77-24,575, 122 pages.

## A COMPARISON OF SELECTED GRADED WORD LISTS AND THE DURRELL ANALYSIS OF READING DIFFICULTY FOR READING LEVEL PLACEMENT

VEON, Betty Ann Spillman, Ed.D.
University of Northern Colorado, 1976

### Purpose of the Study

The purpose of the study was to compare the relationship of reading levels derived from the Durrell Analysis of Reading Difficulty to those obtained from three selected graded word lists: San Diego Quick Assessment, Slosson Oral Reading Test, and the reading subtest of the Wide Range Achievement Test.

### Procedures

The four selected tests were administered and reading level placements for each student for each test were derived. Correlations were computed between the four instruments administered to ninety elementary students in grades one through six. Statistical comparisons were made between the mean reading levels for each of the instruments for the total group, grades one through three, grades four through six, and for the males and females of each group. The discrepancy in grade placement was determined for each student by comparing the results of the Durrell Analysis of Reading Difficulty with each of the word lists. Descriptive comparisons for individual students regarding ranking and instructional group placement were included.

### Conclusions

Based on the findings and subject to the limitations imposed by this study, the following conclusions were drawn: 1. Initial placement in instructional materials using the results of either the selected word lists or the Durrell Analysis may be more justifiable for primary level that upper grade students. 2. No consistent pattern was apparent when comparing students' Durrell Analysis results with the word lists results in the rank placements for students' reading levels and their subsequent reading group assignments.

Results of this study generally supported the concept that the nature and emphases of reading in the primary grades is markedly different from that for grades four through six to warrant the use of different procedures of instruments for assessing reading for the two levels.

Order No. 77-11,088, 175 pages.